Package ‘chngpt’

February 19, 2015

LazyLoad yes
LazyData yes
Version 2014.12-2
Title Change Point Logistic Regression
Author Youyi Fong
Maintainer Youyi Fong <yfong@fhcrc.org>
Depends R (>= 3.0.0), kyotil, MASS
Suggests RUnit, mvtnorm

Imports

Description Hypothesis testing in change point logistic regression with or without interaction terms. Several options are provided for testing in models with interaction, including a maximum of likelihood ratios test that determines p-value through Monte Carlo. Estimation under change point model is also included, but less developed at this point.

License GPL (>= 2)
NeedsCompilation no
Repository CRAN
Date/Publication 2014-12-06 00:17:46

R topics documented:

- chngpt .......................................................... 2
- chngpt.test .................................................. 2
- chngptm ....................................................... 4
- sim.alphas .................................................. 6
- sim.my ....................................................... 6
- sim.pastor .................................................... 7
- sim.sigmoid ................................................... 8

Index 10
_description

Please see the Index link below for a list of available functions. The main testing function is chngpt.test().

usage

chngpt.test (formula.null, formula.chngpt, data, interaction.method=c(
  "lr.mc", "lr.pastor",
  "weighted.two.sided", "weighted.one.sided", "weighted.single.arg",
  "main.itxn", "main.only", "itxn.only"),
  chngpts=NULL, lb.quantile=.1, ub.quantile=.9, chngpts.cnt=50, b=-30,
  single.weight=1,
  mc.n=5e4,
  prob.weights=NULL,
  verbose=FALSE
)

antoch.test (formula, data, chngpt.var, plot.=FALSE)

## S3 method for class 'chngpt.test'
plot(x, ...)

arguments

formula.null formula for the null model. The outcome needs to be a 0/1 variable.
formula.chngpt formula for the change point model. For example, suppose formula.null=y~z and we want to test whether I(x>cutoff) is a significant predictor, formula.chngpt=~x
If instead we are interested in testing the null that neither I(x>cutoff) nor z*I(x>cutoff) is a significant predictor, formula.chngpt=~x*z
data data frame.
interaction.method Only applicable when formula.chngpt contains interaction. See details.
chngpt.test

chngpts A grid of potential change points to maximize over. If not supplied, they will be set to a vector of length chngpts.cnt equally spaced between lb.quantile and ub.quantile.

lb.quantile number. The lower bound in the search for change point in the unit of quantile.
ub.quantile number. The upper bound in the search for change point in the unit of quantile.
chngpts.cnt integer. Number of potential change points to maximize over.
mc.n integer. Number of multivariate normal samples to generate in the Monte Carlo procedure to evaluate p-value.
b. numeric. Slope of logistic function that approximates the step function
single.weight numeric. Only used when interaction.method="weighted.single.arg".
prob.weights numeric. Not yet fully supported.
verbose Boolean.
chngpt.var string. Name of the predictor to detect change point
plot. Boolean. Whether to make a plot.
formula formula.
x An object of type chngpt.test.
... arguments passed to or from methods

Details

When formula.chngpt does not contain interaction term, a test based on the maximum of score statistics is performed. When formula.chngpt contains interaction term, one of the following tests is performed depending on the value of interaction.method. Unless otherwise specified, method for determining p-values is based on Monte Carlo (Fong et al. 2014).

- lr.mc: test statistic is maximum of likelihood ratio statistics.
- lr.pastor: test statistic is maximum of likelihood ratio statistics. Method for determining p-value is based on inequality (pastor.barriuso et al N RPPSI)
- weighted.two.sided: test statistic is maximum of weighted combination of scores that covers both directions of interaction.
- weighted.one.sided: test statistic is maximum of weighted combination of scores that assumes power is enhanced when beta1 and betaR have the same sign.
- weighted.single.arg: test statistic is a weighted combination of scores using weight provided in single.weight.
- main.itxn: test statistic is maximum of two vectors of scores corresponding to the score for the main effect and the score for the interaction effect.
- main.only: test statistic is maximum of one vectors of scores corresponding to the score for the main effect.
- itxn.only: test statistic is maximum of one vectors of scores corresponding to the score for the interaction effect.

If there are missing values in the chngpt formula, those rows will be removed from the whole dataset, including null model and chngpt model.

antoch.test is only implemented for main effect only and is based on Antoch et al. (2004). Also see Fong et al. (2014).

Value

A list of class htest and chngpt.test

- p.value P-value
- chngpts Vector of change points evaluated
- T Standardized absolute score statistics
- V.S.hat Estimated variance-covariance matrix of the score statistics
References


Examples

```r
seed=2
beta=log(.6)
alpha=sim.alphas[["sigmoid2_norm"]]["4.5", paste(log(.6),"",sep="")]
if(is.null(alpha)) stop("alpha not found")
data=sim.sigmoid("sigmoid2", n=250, seed=seed, alpha, beta, x.distr="norm", e.=4.5, b.=-Inf)

test = chngptm.test(formula.null=y~z, formula.chngpt=~x, data)
test
plot(test)

## Not run:
# not run because otherwise the examples take >5s and that is a problem for R CMD check
# has interaction
test = chngptm.test(formula.null=y~z, formula.chngpt=~x*z, data)
test
plot(test)

## End(Not run)
#antoch.test(y-1, dat, "x.star")
```

---

**chngptm**

*Estimate change point logistic model*

**Description**

Estimate change point logistic model

**Usage**

`chngptm(formula.null, formula.chngpt, data, search.all.thresholds=NULL, tol = 1e-04, maxit = 100, chngptm.init=NULL, prob.weights=NULL, verbose=FALSE)`

## S3 method for class 'chngptm'
chngptm

```r
c coef(object, ...)  
## S3 method for class 'chngptm'
vcov(object, ...)  
## S3 method for class 'chngptm'
print(x, ...)  
## S3 method for class 'chngptm'
summary(object, ...)
```

**Arguments**

- `formulaNnull`  formula. Null model formula.
- `formulaNchngpt`  formula. Change point variable formula.
- `data`  data frame.
- `searchNallNthresholds`  Boolean. If NULL, then when sample size is less than 500, it is set to TRUE. If TRUE, search all thresholds.
- `tol`  numeric. Stopping criterion on the coefficient estimate.
- `maxit`  integer. Maximum number of iterations in the outer loop of optimization.
- `chngptNinit`  numeric. Initial value for the change point.
- `probNweights`  numeric vector
- `verbose`  Boolean.
- `x`  chngptm fit object.
- `object`  chngptm fit object.
- `...`  arguments passed to or from methods

**Details**

Please note that the performance of the estimated variance of the coefficient estimates have not been studied thoroughly.

**Value**

A an object of type chngptm with the following components

- `converged`  Boolean
- `coefficients`  vector. Estimated coefficients. The last element, named ".chngpt", is the estimated change point
- `test`  htest. Max score test results
- `iter`  integer. Number of iterations
Examples

data = sim.sigmoid("sigmoid4", n=250, seed=1, alpha=sim.alphas[["sigmoid4_norm"]]["3.4","0"],
    beta=0, x.distr="norm", e.=3.4, b.=-Inf)

fit.1 = chngptm (formula.null=y-z, formula.chngpt=-x, data, tol=1e-4, maxit=1e3)
print(fit.1)
summary(fit.1)

## Not run:
# not run because otherwise the examples take >5s and that is a problem for R CMD check
# has interaction
fit.2 = chngptm (formula.null=y-z, formula.chngpt=-x*z, data, tol=1e-4, maxit=1e3
    , search.all.thresholds=TRUE)
print(fit.2)
summary(fit.2)

## End(Not run)

---

### sim.alphas

#### Simulation Parameters

**Description**

Simulation Parameters

**Usage**

data(sim.alphas)

**Format**

List of 6. Names: sigmoid2_norm, sigmoid2_norm3, sigmoid3_norm, sigmoid3_norm3, sigmoid4_norm, sigmoid4_norm3. Each element is a 5x4 matrix

---

### sim.my

#### Simulate data

**Description**

Simulate data

**Usage**

sim.my(n, seed, label, alpha, beta, e. = NULL, b. = NULL, tr. = NULL)
Arguments

- **n**: Sample size
- **seed**: Seed for random number generator
- **label**: A character string which specifies the simulation scenario. sigmoid4, sigmoidgam4, elbow4
- **alpha**: Regression parameter
- **beta**: Regression parameter
- **e**: Inflection point for the logistic transformation (the log scale)
- **b**: Slope for the logistic transformation
- **tr**: Threshold point

Details

When the label starts with elbow, the transformation on x.star is elbow shaped. When the label starts with sigmoid, the transformation on x.star is sigmoid shaped. Data simulated from logit(Pr(Y==1))=alpha + beta*(transformed x.star).

Value

A data frame with columns: y, x.star, x.star.expit (if label starts with sigmoid), x.star.tr (if label starts with elbow), x.bin.med (x.star dichotomized at median), x.tri (x.star trichotomized at tertiles).

Examples

```r
alpha=-1; beta=log(0.2)
e=5; b=-30; t=1
dat=sim.my(n=250, seed=1, label="sigmoid4", alpha, beta, e=e, b=b)
```

---

*sim.pastor*  
*Simulate data according to one of the scenarios considered in Pastor-Barriuso et al 2003*

Description

Simulate data according to one of the scenarios considered in Pastor-Barriuso et al 2003

Usage

```r
sim.pastor(seed)
```

Arguments

- **seed**: Seed for the random number generator.
**Value**

A data frame with columns: y, x.star, x.star.expit, and x.bin.med (x.star dichotomized at median).

**Examples**

```r
dat = sim.pastor(seed = 1)
```

---

**sim.sigmoid**

*Simulation Function*

**Description**

Simulation Function

**Usage**

```r
sim.sigmoid (label, n, seed, alpha, beta, coef.z = log(1.4), x.distr = "norm", e = NULL, b = NULL)
```

**Arguments**

- `label`: string. Simulate scenario, see details.
- `n`: 
- `seed`: 
- `alpha`: 
- `beta`: 
- `coef.z`: numeric. Coefficient for z.
- `x.distr`: string. Possible values: norm (normal distribution), gam (gamma distribution)
- `e`: 
- `b`: 

**Details**

When label is "sigmoid1", an intercept only model is the data generative model. When label is "sigmoid2", a binary covariate z is also part of the data generative model.
Value

A data frame with following columns:

- y: 0/1 outcome
- x: observed covariate that we are interested in
- x.star: unobserved covariate that underlies x
- z: additional covariate

In addition, columns starting with 'w' are covariates that we also adjust in the model; columns starting with 'x' are covariates derived from x.
Index

*Topic distribution
  chngpt, 2

antoch.test(chngpt.test), 2

chngpt, 2
chngpt.test, 2
chngptm, 4
coef.chngptm(chngptm), 4

plot.chngpt.test(chngpt.test), 2
print.chngptm(chngptm), 4

sim.alphas, 6
sim.my, 6
sim.pastor, 7
sim.sigmoid, 8
summary.chngptm(chngptm), 4

vcov.chngptm(chngptm), 4